

Genetic Algorithm based Threshold Sensitive Routing Protocol for Wireless Sensor Network

Anjali Priyanka Tigga

Roll Number 213CS1153

under the supervision of

Dr. Manmath Narayan Sahoo



Department of Computer Science and Engineering
National Institute of Technology
Rourkela-769008, India

Genetic Algorithm based Threshold Sensitive Routing Protocol for Wireless Sensor Network

Dissertation submitted on

MAY 27, 2015

to the department of

Computer Science and Engineering

of

National Institute of Technology, Rourkela

in partial fulfillment of the requirements

for the degree of

Master of Technology

by

Anjali Priyanka Tigga

(Roll No. 213CS1153)

under the supervision of

Dr. Manmath Narayan Sahoo

Department of Computer Science and Engineering

National Institute of Technology

Rourkela - 769008, India



Computer Science and Engineering
National Institute of Technology Rourkela
Rourkela - 769008, India www.nitrkl.ac.in

May 27, 2015

Certificate

This is to certify that the work in the thesis entitled **Genetic Algorithm based Threshold Sensitive Routing Protocol for Wireless Sensor Network** by *Anjali Priyanka Tigga*, bearing Roll Number 213CS1153, is a record of an original research work carried out by her under my supervision and guidance in partial fulfillment of the requirements for the award of the degree of *Master of Technology* in *Computer Science and Engineering Department*. Neither this thesis nor any part of it has been submitted for any degree or academic award elsewhere.

Dr. Manmath Narayan Sahoo

Assistant Professor

Department of CSE

NIT, Rourkela

Acknowledgment

First of all, I would acknowledge my supervisor Dr. Manmath Narayan Sahoo, who has been the driving force behind this work. I sincerely, thank him for introducing me to the field of Routing Protocols in Wireless Sensor Networks and giving me an opportunity to work under him. His immense faith in this topic and the patience shown towards people trying to perceive their analytical and practical skills required in a research scholar has been a great motivation. His faith has led to the completion of the thesis work. I regard it my good fortune to have an opportunity to work with such a knowledgeable and wonderful person.

I would like to thank Prof. Suchismita Chinara, who during the summer took the step to conduct seminars relating to emerging wireless sensor networks. It made a huge impact to decide the field to work upon in the academic year. Also had an opportunity to comprehend WSNs importance as in nearby Rourkela Steel Plant (RSP) itself.

I would like to exhibit my sense of obligation towards the Head of Department of Computer Science and Engineering Department, Prof. Shantanu Kumar Rath, who has always urged to discover a researcher's discipline in oneself. He has also ensured this through various conducts on regular basis for which I am very grateful. I thank him for bringing the best out of a student like me.

I would also like to thank Prof. Bidyut Kumar Patra, who encouraged me during the presentation ascertaining that the base work taken for thesis is from recognized publications and authors. And work done is also remarkable.

I would also like to thank all my faculty members, secretarial staff, PhD scholars, my seniors, juniors and all my acquaintances to pour in their valuable suggestions, cooperation and stimulations during the entire period of work.

At last but not the least I heartily thank my family who never doubted on my efficiency to achieve this long way to work and learn as a research scholar.

Anjali Priyanka Tigga

Abstract

Wireless sensor network (WSN) is an advancing technological field of wireless communication with a wide range of applications. It comprises of geographically wide-spread unattended micro-sensors. Such autonomous characteristics of nodes have led to design of various routing protocols which encounter performance augmentation with respect to energy efficiency and network lifetime. Design of routing protocols is highly contingent on the application which uses it. However, hierarchical cluster-based structure is ascertained to yield considerable potential. Rooted from it is the widely known cluster-based LEACH protocol. Considerable contribution is also done in protocols based upon genetic algorithms such as ERP and HCR. They tend to extend network lifetime but are liable to attainable modifications which can direct to better performances. The aim of this paper is to acquire effective fitness function to construct clusters efficiently which can operate with high stability and attain low energy dissipation over the rounds of transmission. Another moderation is the route configuration for aggregated data to be transmitted along the cluster heads only. Simulation and comparison with LEACH, TEEN and ERP shows that GA-TSRP is significantly stable for a number of transmission rounds. And thus, energy utilized is spread uniformly over an adequate duration of sensor network lifetime.

Keywords: Genetic algorithm, hierarchical clustering, stability period, energy efficiency, data aggregation.

Contents

Certificate	ii
Acknowledgement	iii
Abstract	iv
List of Figures	vii
List of Tables	viii
1 Introduction	1
1.1 Components of Sensor Nodes	3
1.2 WSN Applications	5
1.3 Design challenges for routing in WSN	7
1.3.1 Node Deployment	7
1.3.2 Energy Consumption	7
1.3.3 Data Reporting	8
1.3.4 Node Heterogeneity	8
1.3.5 Fault Tolerance	8
1.3.6 Scalability	9
1.3.7 Network Dynamics	9
1.3.8 Transmission Media	9
1.3.9 Area Coverage	10
1.3.10 Aggregation of Data	10
1.3.11 Quality of Service (QoS)	10
1.4 Motivation	10
1.5 Objective of Research	11

1.6	Summary	11
2	Literature Review	12
2.1	Classification of sensor network routing protocols	12
2.2	Related Work	13
2.2.1	SPIN (Sensor Protocols for Information via Negotiation) . . .	13
2.2.2	Directed Diffusion	13
2.2.3	Cougar	13
2.2.4	Rumor Routing	14
2.2.5	TTDD (Two-Tier Data Dissemination)	14
2.2.6	Energy Aware Routing	15
2.2.7	LEACH (Low-Energy Adaptive Clustering Hierarchy)	16
2.2.8	Pegasis	17
2.2.9	TEEN (Threshold sensitive Energy Efficient sensor Network protocol)	18
2.2.10	APTEEN (Adaptive Periodic Threshold-sensitive Energy Efficient Sensor Network Protocol)	19
2.2.11	SEP (Stable Election Protocol)	19
2.2.12	Genetic Algorithm based Routing Protocols	20
2.3	Summary	21
3	Proposed Methodology	22
3.1	Background	22
3.2	Proposed Protocol	23
3.2.1	Energy model	26
3.2.2	Assumptions	26
3.2.3	Phases	27
3.3	Summary	29
4	Simulation Result	30
4.1	Analysis of GA-TSRP	30
4.2	Summary	34
5	Conclusion and Future Work	35
	Bibliography	36

List of Figures

1.1	Depiction of Wireless Sensor Network	2
1.2	Sensor Node Architecture	4
2.1	Time-line of LEACH	17
2.2	Operation Flow Chart of TEEN	18
2.3	Hierarchical Clustering Scheme	19
3.1	Path Formation Phase	28
4.1	Comparison based on nodes alive in network over the rounds	32
4.2	Comparison based on energy left in network over the rounds	32
4.3	Comparison based on cumulative energy dissipation over the rounds	33

List of Tables

2.1	Comparison of flat routing protocols	15
2.2	Comparison of hierarchical routing protocols	20
4.1	Parameter settings for simulation	31

Chapter 1

Introduction

The present-day advances in micro-sensor technology provides with great opportunities for a wide range of applications. These extend across the fields of monitoring health, traffic, weather and toxins, surveillance of fields, observance of global resources and geological activities, industrial applications etc. Prominent research work has been done to address data aggregation schemes, routing techniques and efficient energy usage in WSNs. Since, the sensor nodes have energy and bandwidth constraints, there is an eminent need for novel techniques that can lengthen network lifetime and eliminate energy inefficiency. The major functionality of a WSN is to observe and report about a predefined phenomenon in a physical environment and transmit the sensed data to a central node called a base station (BS). Unattended adversely deployed micro-sensors need to be fault-tolerant and efficient enough to sustain longer. The concept of routing comes into existence when direct transmission is not feasible. Thus, there exist a number of schemes which proclaims efficient usage of energy while route formation and selection of sensor nodes for direct or multi-hop transmission to the base station. Routing in WSNs is far more challenging than wired due to absence of global addressing scheme, presence of numerous nodes and corresponding data traffic which is highly redundant. Also, transmission, processing, data storage and battery power is constrained limiting the schemes to be more coherent. Figure 1.1 shows a diagrammatic representation of

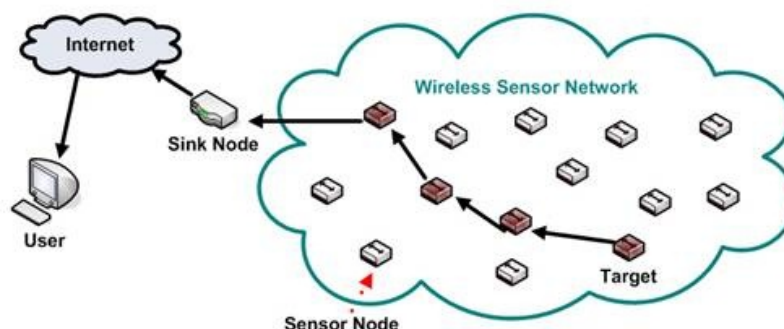


Figure 1.1: Depiction of Wireless Sensor Network

the wireless sensor network.

Due to such limitations many algorithms and routing schemes have been proposed. On the basis of network structure routing protocols for WSN can be classified as flat network routing, hierarchical network routing and location-based network routing [1]. In flat network routing every node has equal functionality and thus, sensing and routing tasks are worked upon together. Contrary is the case with hierarchical network routing in which roles are divided among cluster heads and the member nodes. In location network routing route is worked upon by the sensor node's position. Again these protocols can be classified based upon protocol operation they perform as multi path-based, query-based, negotiation-based, QoS-based, or coherent-based routing techniques [2]. LEACH is one of the most famous hierarchical cluster -based network routing protocol which is accepted widely due to energy efficiency and simplicity to construct the clusters [3]. Also it is induced that hierarchical clustering structure is likely to provide with scalable and energy-efficient results. In the clustering scheme sensed data is propagated from normal member nodes to cluster heads which on aggregation sent to the BS. These are proficient enough than any other scheme and in general, are heuristic-based aiming to generate minimum clusters and minimum transmission distance. Many of these techniques work on the principle of "equality". Thus, every node gets an equal chance to operate as a cluster-head without taking into

consideration whether after being the CH it will sustain for next rounds or not.

Biological systems exhibit scalability, robustness, resistance to individual failures thereby following the principle of “improvement” which considers every single individuals capacity. Hence, recently evolutionary algorithms (EAs) have attained researcher’s attention so as to develop hierarchical cluster-based routing protocols for WSNs. Primary goal of these genetic algorithm based cluster routing protocols is to form clusters dynamically so that energy consumption in the network is diminished thus, extending network lifetime. In genetic algorithm based routing like GA, HCR and ERP, the location of nodes exploited and cluster formation done as selection through set of computed solutions or chromosomes [4], [5], [6]. They have already proved their efficiency when compared to LEACH. However, they are still liable to remarkable modifications which can make them more stable. Stability is said to be attained for an interval by the network when there is no dead node present in the network. After this period network nodes tend to die in a frequent manner. The work done in this paper tries to attain more stability and network lifetime by introducing transmission distance and residual energy as the main factors for the fitness function used in EAs. Further, final path is also formulated which tends to minimize the burden of direct or multi-hop transmission by the cluster heads directly to BS.

This thesis work demonstrates the influence of factors stated in formulating set of solutions through EAs selection, crossover, mutation and fitness evaluation. And proves the liability of proposed methodology through comparison of network life and energy dissipation with protocols namely, LEACH, TEEN and ERP.

1.1 Components of Sensor Nodes

The small sized sensor node comprises of a lot of components which makes it possible efficiently to sense, compute and transfer the data over the wireless sensor network. Primarily it consists of the following:

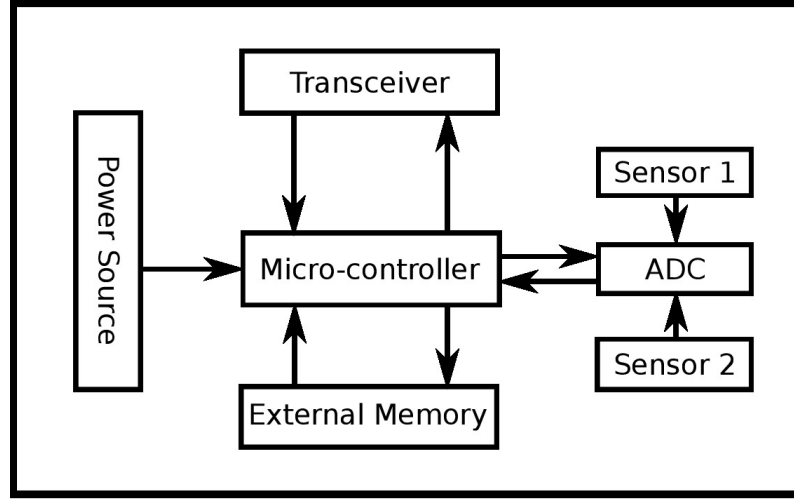


Figure 1.2: Sensor Node Architecture

- (a) **Micro Controller:** It is a general purpose optimized processor for embedded applications with special care taken for low power consumption. Its function is to process data, execute time sensitive tasks, control the communication protocols and other components. Active, idle, listen, sleep are among the various modes of operation of a controller.
- (b) **Transceiver:** Transmitter and receiver are the communication devices used in a sensor node. Transmitter and receiver integrated together in a single device is known as a transceiver. The transmission channel used are radio frequency (RF), infrared and optical communication. As the lasers (optical communication) need line-of-sight to communicate and infrared's broadcasting capacity is limited, WSN applications make use of radio frequency based communication. Its operational state comprise of transmit, receive, sleep and idle.
- (c) **External Memory:** In general, on-chip flash memory is used due to their low cost and storage capacity. However, requirements of memory is application dependent. The memory comprise of two categories of memory namely, user memory which contains application related or personal data and program

memory. Program memory is used for programming the device and may also contain devices identification data.

- (d) **Power Source:** Sensor nodes consume energy while sensing, communicating and processing data. But, more energy is required while communication of data. Power storage is done in batteries or capacitors. A wide variety of sensor nodes try to scavenge energy from the nearby environment.
- (e) **Sensors:** A sensor node may comprise of one or more small size, low energy, adaptive sensors. Sensor is a hardware device which produces a quantifiable, measurable response when a physical phenomenon like temperature, pressure being sensed changes. Thus, they measure the physical data of the parameter being monitored. The signal produced is analog which is digitized by the ADC (analog-to-digital converter) and then propagated to the micro controller for further processing. They are classified into three categories namely, passive omni-directional sensors, narrow-beam sensors and active sensors.

1.2 WSN Applications

The spatially distributed sensors in an area has a wide range of applications classified as process management, environment sensing, industrial monitoring etc. Given below are some of the many applications of WSNs.

- (a) **Area observing:** Area observing is a basic use of wireless sensor network. In area observing, the WSN is deployed over an area where some phenomenon like temperature, pressure etc is to be observed. Such as in military sensors are utilized to detect any adverse interruption.
- (b) **Monitoring health care:** The two kinds of medical applications are wearable and embedded. In wearable, devices are utilized on the body surface of a human or positioned closely to the client. Whereas, the implantable devices

are embedded inside a living body. Other applications include estimation of body position of an individual, general checking of patients at hospitals and homes. Body-area systems can monitor and record data about a being's fitness, energy used etc.

- (c) **Prevention of natural disaster:** Remote sensor networks can predict outcome from a natural calamity thereby communicating sensed parameter to the main station. Many waterways comprise of remote hubs to observe changes in water levels and report when level is beyond safe measure.
- (d) **Avalanche identifier:** An avalanche identification system makes use of a remote sensor network system to identify the changes in soil with respect to different parameters that may happen during or before an avalanche. Through the gathered information it might be possible to detect oncoming avalanches before it happens.
- (e) **Data logging:** Wireless sensor networks are used for gathering information while observance of environment such as checking the temperature, pressure, heat etc. The facts thus collected can then be used to show how the functioning frameworks. WSNs provide with "live" information contrary to the customary logging systems.

1.3 Design challenges for routing in WSN

WSNs comprise of various restrictions such as confined power supply, confined computation power and confined bandwidth. One of the main design goals of WSNs is to ensure prolongation of network lifetime and carry out data communication and prevent abrupt energy depletion of the network. Following are some of the routing challenges and design issues that affect routing process in WSNs.

1.3.1 Node Deployment

Node deployment in WSNs depends upon application and thus, affects the routing protocols performance. The arrangement can be either deterministic or randomized. In deterministic, the sensors are placed manually and data routed through pre-defined paths. Nonetheless, in irregular organization, the nodes are scattered haphazardly making a framework in a specially appointed way. In the event that the resultant distribution of nodes is not uniform, ideal grouping gets to be important to permit integration and empower vitality effective network operation. Communication among sensors is normally inside short transmission because of energy and data transfer capacity restrictions. Along these lines, it is in all probability that a route will comprise of different remote hops.

1.3.2 Energy Consumption

While computing sensed data and transmitting relevant information sensor nodes deplete considerable amount of limited energy it has as stored in the battery. So, computation and communication in need to conserve energy as much as possible. Lifetime of a sensor node depends on its battery lifetime. Also, in multi-hop routed WSN every node has a dual role as sender and router. Sensor nodes can be defected due to power failure and can cause change in topology. In such condition packets need to be rerouted after the network is reorganized.

1.3.3 Data Reporting

Data to be sensed and reported in WSNs depends upon the application and its need for time critical reporting of data. Data reporting can be classified into event, time, query driven and hybrid. In event-driven model data is reported immediately when there occurs an event described a priori. Query driven model reacts accordingly to the query sent by BS. For applications which need periodic data reporting time-driven model is used. For this to happen, sensor nodes are not on all the while. Instead, they operate only at regular intervals sensing and communicating. Also combination of any model described is possible. Data reporting model influences routing technique used with respect to energy dissipation and the stability of route.

1.3.4 Node Heterogeneity

In many cases all sensor nodes are assumed to be homogeneous with equal capacity to compute, communicate and contain equal power. However, homogeneity or heterogeneity depends on the application for which sensor network is deployed. Heterogeneity present in network can pose many technical issues in routing. This leads from the fact that heterogeneous sensor node network can have multiple data reporting models and varying quality of service. In the hierarchical protocols cluster-head node's role is dissimilar from the cluster-member normal sensors. The cluster-heads depending on clustering scheme can be selected from the nodes deployed or can have more power, bandwidth etc from the time of network initialization itself. So the transmission load to BS is managed by cluster heads.

1.3.5 Fault Tolerance

Sensors can fail due to some environmental damage, power loss etc. However, failure of a sensor node must not hinder the network from its task. On failure of large number of nodes, routing and media access control (MAC) protocols need to form new route to BS. This can be achieved by adjustment of transmitter's power and

rate of signals so as to reduce energy dissipation and rerouting of packets at areas where energy is comparatively more. Thus, for a fault network various levels of redundancy may be needed.

1.3.6 Scalability

If one wants to add or remove sensor nodes from the field and have the operation going on it should be possible in the WSN. For this to be possible routing protocols should also be scalable. It can be achieved by number of nodes not transmitting every while

1.3.7 Network Dynamics

Contrary to most network architectures which assume that sensor nodes are stationary, depending on application need there could be mobile base stations and mobile sensor nodes. Message routing among mobile nodes is rather a difficult task because stability is not present with respect to routing, energy and bandwidth. Also, the sensed phenomenon can be dynamic too as in tracking system or static as in environment temperature monitoring. Static phenomena could be monitored with reactive network. But dynamic phenomena need to have a network such that data reported periodically creating traffic while routing.

1.3.8 Transmission Media

Multi-hop sensor nodes in WSNs are linked for communication through wireless channel. Error rate of messages and fading of signals also affects the network operation. Generally 1-100 kb/s bandwidth is required in WSNs. MAC is designed accordingly as the transmission media used. One of the most used approach is the TDMA based protocols which sustain more energy compared to the CSMA, contention based protocol. Another good approach is to use bluetooth technology.

1.3.9 Area Coverage

Sensors can sense to a particular range. So the depiction of field sensed vary according to range and accuracy attained by nodes. Accordingly, area coverage should be achieved efficiently.

1.3.10 Aggregation of Data

As nodes are deployed densely in a field many nodes transmit similar set of data increasing redundancy. To eradicate this data aggregation is used. Combining data regarding an event from different sources is termed as data aggregation. This technique makes the routing protocols energy efficient and also optimizes the data transfer. Signal processing methods can be used to aggregate data which is termed as data fusion. Beamforming used in signal processing is used to combine the incoming signals and reduce the noise.

1.3.11 Quality of Service (QoS)

QoS comes into play with applications which need timely deliverance of data. For that to happen, data needs to be delivered under a bounded latency. In many applications low network lifetime is prior challenge to overcome than to attain quality of data. As energy gets depleted quality of data packets decrease. Therefore, energy-aware routing protocols have to address this need also.

1.4 Motivation

A sensor network is said to be accurate and reliable only when sensed data forwarded comes from large number of member nodes aggregated data. This can happen only when nodes do not die abruptly making the network unreliable. And the battery power decides how long network will be alive. Thus, death of first node decides stability of the network. Presently, protocols stay alive for long time but after

certain point of time degrade steeply. The degradation starts after the death of first node in the network. This fact led to work upon factors which can reduce early death of nodes over time.

1.5 Objective of Research

For longevity with validity of data collected energy dissipation over node death should be minimized. Prior objective is to formulate a cluster formation method through which sensor network can dissipate energy substantially. Substantial energy dissipation means that the network should prolong for a long time with the nodes alive in great number. And also route final aggregated data in such a way that energy consumption is minimized.

1.6 Summary

This chapter introduced about the trending wireless sensor networks with a description regarding sensor node architecture, applications of WSN and the design challenges of routing in WSN. The next chapter overviews previous work done in the field of WSN routing protocols.

Chapter 2

Literature Review

This chapter gives a brief description of various categories of sensor network routing protocols.

2.1 Classification of sensor network routing protocols

Based upon network structure classification can be done as flat-based routing, hierarchical-based routing and location-based routing [2]. In flat-based routing every node has equal function to handle. Whereas, nodes in hierarchical-based routing possess variant functionalities. As the name suggests, location-based routing use the location information of nodes to transmit data packets over the network. Depending upon protocol operation classification can be done as multi-path-based, query-based, negotiation-based, QoS-based, or coherent-based routing techniques [1]. In multi-path-based routing technique redundant data transmitted through multiple paths to the final destination. In query-based routing technique event data is transmitted only when the BS queries about it to the network. In negotiation-based routing technique redundancy of data removed by transmitting only after negotiation with other nearby nodes ensuring the necessity of the data

forwarded.

2.2 Related Work

2.2.1 SPIN (Sensor Protocols for Information via Negotiation)

SPIN is an adaptive protocol which eradicates the deficiencies like implosion, overlap and resource blindness faced by classic flooding [7]. SPIN provides with negotiation and resource-adaptation which together overcomes above deficiencies. Negotiation helps to eradicate implosion and overlap problem by negotiating before transmitting data. For this, data descriptor called meta-data is used to describe the data. Meta-data transmitted before actual data transmission is small in size. Each sensor node also comprise of resource manager to keeps track of resource consumption; before transmitting or processing data the manager is probed about it. Thus, confirmed that certain activities not done when energy is low.

2.2.2 Directed Diffusion

Directed diffusion is designed for a simple remote-surveillance network which disseminates sensor query and then, processing is done. They have focused in the design of dissemination techniques for tasks and events. Data at nodes is named by attribute- value pair. In directed diffusion dissemination is data-centric, best path formation is done through reinforcement-based adaptation and data aggregated in network itself [8].

2.2.3 Cougar

The authors have stated that the existing sensor networks assume sensors to be preprogrammed for a particular function and transmit data to a centralized front-end

where further data aggregation, analysis and storage is done [9]. This approach lacks changing the system behavior as when needed and also, trades costly communication for local computation which are cheap. Thus, they have introduced the Cougar approach to tasking sensor networks through declarative queries which does not resemble preprogrammed sensors. And thus, have also introduced sensor network as a processing platform. Its approach is driven by two main factors i.e., declarative queries and in-network data aggregation and processing. In declarative queries users and application programs issued queries are processed remains abstracted from the network layer functions. This is attained by a query proxy which lies between the network layer and application layer and interacts with both the layers present at every nodes query layer. Reduced energy consumption acquired by in-network data aggregation and processing by the query proxy at each node.

2.2.4 Rumor Routing

In this method nodes on observing a particular event describe and analyze and route queries. To attain this, there exist long-living packets termed *agents*. Events table exist which gets updated by node locally on detection of an event and thus, generates an agent. Distant nodes are informed about local events through agents [10]. The nodes that know the route can respond by looking-up their event table when a query is generated. Hence, there is no need to flood entire network reducing the communication cost. It supports only a single path between the source and destination. If events are large and interests are not enough it would be infeasible to maintain agents and event-tables at every node.

2.2.5 TTDD (Two-Tier Data Dissemination)

This approach uses grid formed among nodes to transmit data to multiple base stations. Grid is formed on occurrence of an event by nearby node sensing the data along path such that each node forwards to nearest signal strength neighbor storing

Table 2.1: Comparison of flat routing protocols

Author, Year	Key Points	Remark
W. R. Heizelman et al, 1999 [7]	<ul style="list-style-type: none"> • Data naming through meta-data • ADV-REQ- DATA, event driven, multi-hop • Negotiation and resource based 	<ul style="list-style-type: none"> • No guarantee of data deliverance (interest too away from event generation point)
C. Intanagonwiwat et al, 2000 [8]	<ul style="list-style-type: none"> • Data-centric • Periodic interest sent through sink • Gradient specifies data rate and direction to send events 	<ul style="list-style-type: none"> • Overhead due to data and query matching
Y. Yao et al, 2002 [13]	<ul style="list-style-type: none"> • Declarative queries • In-network data aggregation • Processing using query layer 	<ul style="list-style-type: none"> • Extra overhead due to query layer • Leader nodes need dynamic maintenance
Braginsky et al, 2002 [10]	<ul style="list-style-type: none"> • Variation of [8] • Distant nodes receive information regarding local events through agents. 	<ul style="list-style-type: none"> • Unsuitable for large event numbers • Next hop selection suffers due to heuristic route of event

source information all along until the edge of network. From that point extreme node communicates to BS. Thus, grid formed and BS forwards query to nearest dissemination point. And through grid query finally reaches source which sends back according to query received. On change of network topology there can occur abundant overhead with respect to the recalculation of grid [11].

2.2.6 Energy Aware Routing

This protocol is a reactive protocol based on initiation through destination [12]. It is similar to [8] except that it maintains a set of path instead of single optimal path. Path chosen using probability function whose value depends on how much a paths energy consumption is minimized. It collects information of node location so as to build up an addressing scheme and thus, inefficient compared to directed diffusion.

2.2.7 LEACH (Low-Energy Adaptive Clustering Hierarchy)

In LEACH, authors have considered that the sensor nodes pertain equal functionalities and possess equal initial energy. BS is located far away from the network. It is a hierarchical cluster-based protocol in which cluster selection and formation is done in the network by nodes itself i.e., self-organized [3]. Energy model used is the first order radio model. Clusters are self-organized with usage of a probability function as in Equation (2.1). Nodes get equal chance to become a cluster head based on choosing a probability value between 0 to 1. Predefined number of nodes made cluster head which choose probability value less than threshold value get the chance to become cluster head. CH advertise for cluster member join to nearby nodes and nodes confirm back to CHs through CSMA MAC protocol. Normal nodes present in every cluster transmits to cluster head periodically according to the TDMA schedule broadcast to them by CH. Cluster heads aggregate the received data and send to the BS. Different CSMA codes used by CHs to reduce any interference occurrence due to presence of clusters in a region. Nodes on becoming CH choose randomly from a list of spreading codes. Figure 2.1 depicts the time-line of LEACH protocol.

$$T(n) = \begin{cases} p/(1 - p * (r \bmod (1/p))) & \text{if node } i \in G \\ 0 & \text{otherwise} \end{cases} \quad (2.1)$$

where r is the round number, p is the predefined desired percentage of CHs normally taken to be 5% and G is the set of nodes not been cluster head in the last $(1/p)$ rounds.

However, randomized rotation cannot ensure optimal distribution of CHs since, residual energy is not considered both nodes with low-energy and high-energy have equal probability to become a cluster head. This leads to early and fast declination of the network nodes.

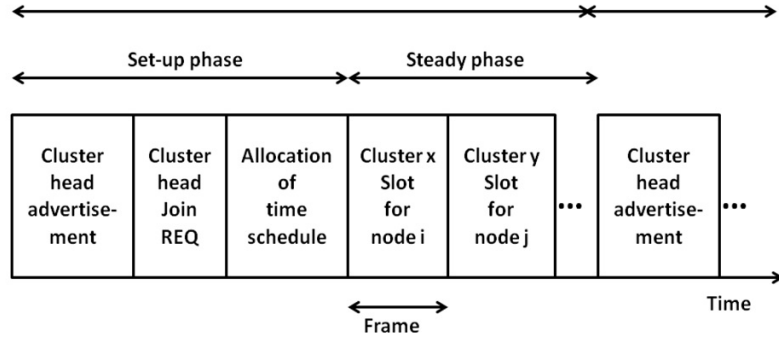


Figure 2.1: Time-line of LEACH

2.2.8 Pegasus

The authors have proposed an improvement over Leach protocol named, Pegasus. Comprises of three phases namely, chain formation, leader selection and data transmission [14]. A single node has the responsibility to send the fused data to BS every round (along chain) which is chosen randomly every round. On select From all ends of the network data received in a chain. Each node between the chain or chains receive and send data. Thus, nodes communicate or disseminate only to neighbor after fusing with ones own data. Chain is made among sensors through greedy approach and not like traveling salesman problem because it is intractable. This is done keeping in mind the radio communication energy parameters. First order radio model as in Leach used here. For construction of chain all nodes have global knowledge about the network. Pegasus considers that every round any single node will transmit to BS but practically, multi-hop communication comes in far away node scenario. Also, there can be an excessive delay when chain head is at the far end of network.

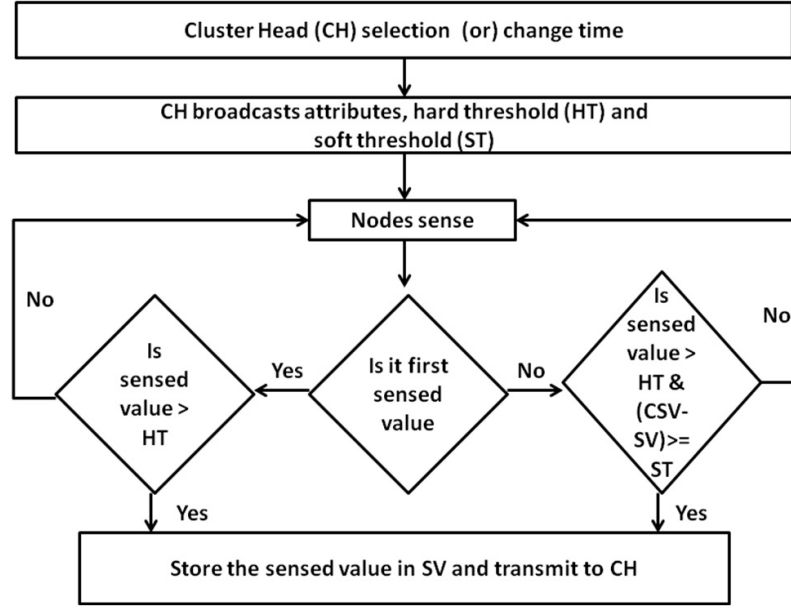


Figure 2.2: Operation Flow Chart of TEEN

2.2.9 TEEN (Threshold sensitive Energy Efficient sensor Network protocol)

The authors have introduced TEEN for reactive networks. Re-active networks are the one which immediately respond to relevant changes in the parameters of interest [15]. They have evaluated it for simple temperature sensing application showing that it out-performs existing conventional sensor network protocols in terms of energy efficiency. Main features of this architecture are: save energy by transmitting to immediate cluster-head, additional computations on data need to be performed only by the cluster head which send data over correspondingly larger distances, for even distribution of energy consumption every node takes turn becoming the cluster-head. Nodes become cluster-head sustain as head for a time interval T (cluster period). Figure 2.2 is an operation flowchart of TEEN. Figure 2.3 shows the hierarchical clustering scheme used.

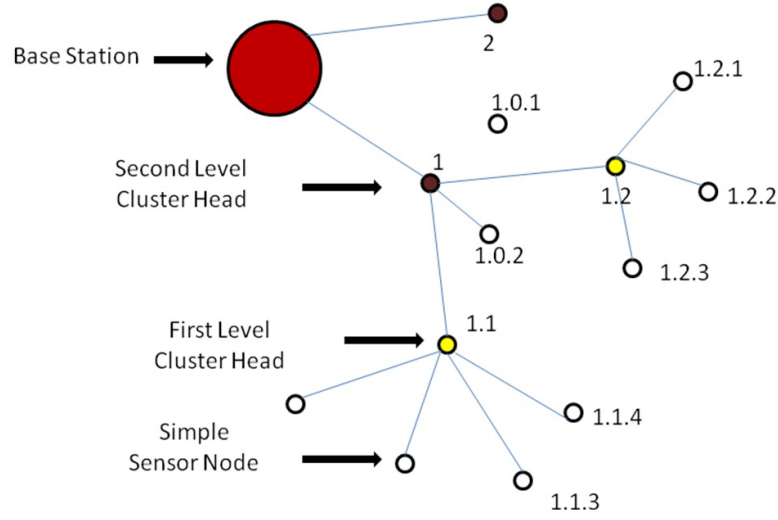


Figure 2.3: Hierarchical Clustering Scheme

2.2.10 APTEEN (Adaptive Periodic Threshold-sensitive Energy Efficient Sensor Network Protocol)

The authors have proposed to create a hybrid network which combines the best features of proactive and reactive networks. It is an extension of [15] which did not had any proactive response. It sends data periodically and also, reacts to sudden changes in values of attribute [16]. Also, evaluated its performance and showed that it is better than other similar protocols. For various query types queried by user, an efficient query handling mechanism is developed. Queries can be handled either by use of centralized system, known as the warehousing approach. Drawback is that the critical data has to be extracted from the database. Another method is to collect data on demand which on other hand will pose delay for the queries concerning time critical data.

2.2.11 SEP (Stable Election Protocol)

In hierarchical cluster based protocols mentioned above nodes are homogeneous. However, in SEP heterogeneous nodes are present i.e., every node does not possess

Table 2.2: Comparison of hierarchical routing protocols

Author, Year	Key Points	Remark
W. R. Heizelman et al, 2000 [3]	<ul style="list-style-type: none"> • Autonomous clusters of nodes • Periodic election of CH • Data to CH using TDMA schedule • CH aggregates data • Multi-hop communication 	<ul style="list-style-type: none"> • Uniform distribution of CHs not guaranteed • Does not account non-uniform energy of node
S. Lindsey et al, 2002 [14]	<ul style="list-style-type: none"> • Communication chain created (Greedy Algorithm) • Sensor node sends its data to its nearest neighbor • Chain head selection random and transmit to sink 	<ul style="list-style-type: none"> • Dynamic topology needed can be an overhead • Method to obtain node location is unspecified
A. Manjeshwar et al, 2001 [15]	<ul style="list-style-type: none"> • Reactive • Transmits only when HT and ST satisfied 	<ul style="list-style-type: none"> • Multiple level cluster formation adds to complexity
A. Manjeshwar et al, 2002 [10]	<ul style="list-style-type: none"> • Advancement of [15] • Support periodic data collection and time-critical situations 	<ul style="list-style-type: none"> • Method to implement threshold based function is an overhead

equal amount of energy initially. SEP is based on weighted election probabilities of each node to become cluster head according to the remaining energy in each node [17]. Further, [18] and [19] are the modifications of SEP.

2.2.12 Genetic Algorithm based Routing Protocols

Genetic Algorithm makes use of prior knowledge about the network to determine best fit solution (i.e., clusters in network). This information can include energy, distance, cluster size etc. In hierarchical cluster-based routing protocol (HCR) nodes self-organize as clusters with each cluster managed by set of associates termed as head-set [6]. CHs elected from head-set through round-robin technique. HCR is then improved using Genetic Algorithm (GA) which determines clusters, their heads, members and the schedule to transmit data [20]. GA is used at the BS and consists

of an optimizer which selects the best solution based on fitness function. Fitness function is based upon parameters such as energy consumption, size of cluster, distance to sink, distance among cluster members.

Evolutionary based routing protocol is also a GA based routing protocol in which fitness function comprise of clusters compaction and separation together with number of clusters [5].

2.3 Summary

This chapter described about the various paradigms used to classify sensor network routing protocols and based on them the routing protocols designed till date. Chapter 3 carries on with the proposed methodology.

Chapter 3

Proposed Methodology

3.1 Background

In genetic algorithm based energy efficient clustering each node in the network is represented in a chromosome by a bit. 1s represent head nodes and 0s represent normal nodes. -1 denote dead nodes. The basic framework of GA is used as given in Algorithm 1. Population consists of a number of individuals called chromosomes [4]. This set of population define the solution to the problem, here clustering. Initialization of population done randomly. After that, new generations evolve through selection, crossover, mutation and test based on fitness for a predefined number of times. The fitness function quantitatively defines the "enhancement". The selection process decides the chromosomes which will crossover and create new chromosomes. Crossover is a binary genetic operator which acts upon two parents [4]. These include single-point, two-point, multi-point, arithmetic operator. It is done after the selection process depending upon the rate of crossover which in general, is around 80 to 95 percent. Chromosomes with better fitness value are selected to crossover. Offspring thus produced join the population already present either replacing the chromosomes which crossover or adding to the population. Through crossover the generation evolved will only have the characteristics of the parents

and no new genetic trait is introduced to the offspring. Therefore, mutation is done at a certain rate in general, 0.5 to 1 percent, adding new genetic traits to the population. Unlike crossover, mutation is a genetic unary operator which at a time acts upon a single chromosome. Bit gets flipped while mutation of a chromosome. Thus pertaining to GA, nodes in a WSN are represented as bits of a chromosome. The fitness of a chromosome evaluated regarding various parameters such as the density of node and energy dissipated.

Algorithm 1: Framework of Genetic Algorithm Used

Data: Fitness Function

Result: Population of network topology

while *maxIteration to produce a generation is not reached* **do**

 select random chromosomes in a group of two;

if *parents probability* \leq *crossover probability* **then**

 crossover the chromosomes;

if *chromosomes probability* \leq *mutation probability* **then**

 mutate the offspring chromosomes;

 compute the fitness of all the offspring;

3.2 Proposed Protocol

The main factor determining the performance of any GA based technique is the choice of a fitness function which is highly correlated to the problem. Selection is done through this function itself. With respect to the WSN, fitness function is formulated so that the energy dissipation is minimized and lifetime of network is increased. The previous GA based protocols [5] and [4] are surely better than the LEACH protocol. But they do not consider the residual energy of nodes present for formulating the fitness function and the end cluster head energy consumption to

transmit packet to the BS. This leads to a network with less stability period. Stress is put on stability period because it is observed in various routing methods that node death graph changes steeply sometimes after the stability period is over i.e., after the first node dies. Also, after this point reliability of data cannot be sured since, nodes die abruptly until all nodes are dead over a period of rounds.

Hence, we have redefined the fitness function to be used in the GA based hierarchical clustering. It is a protocol based on the genetic algorithm and the threshold sensitiveness. Threshold sensitive means the network nodes transmit data only when there is a change based upon given threshold. To efficiently design clustering solutions through GA, we need to consider factors namely transmission distance, average minimum distance, energy left per node, cluster head transmission energy and cluster head count. These factors constitute the fitness function.

- (a) **Transmission distance (TD):** It is the total distance a packet covers from the point of formation to the point where it gets aggregated i.e., CH. This factor needs to be minimized. Equation (3.1) gives the transmission distance of normal member nodes.

$$TD = \sum_{i=1}^{CHs} \sum_{\forall mn \in Cl_i} dis(mn, CH_i) \quad (3.1)$$

- (b) **Average minimum distance (AMD):** It is the average of distance among all the cluster heads in a network. Distance evaluated as the Euclidean distance between cluster heads. This needs to be maximized conforming that the clusters formed are not very close to each other. Equation (3.2) gives the average minimum distance of cluster nodes. nc denotes the number of cluster head.

$$AMD = \sum_{\forall Cl_i, Cl_j, Cl_i \neq Cl_j} dis(CH_i, CH_j) / nc \quad (3.2)$$

- (c) **Energy left per node (EL):** It is the total residual energy of every node present in the network. This factor needs to be maximized conforming that the energy per node is dissipated uniformly over the network so as to increase the stability period. Equation (3.3) gives the summation of energy left per node. E_{oi} denotes the initial energy of node i and ED_i denote the energy dissipated by node i .

$$EL = \sum_{i=1}^n (E_{oi} - ED_i) \quad (3.3)$$

- (d) **Transmission energy dissipated (TED):** It is the sum of energy dissipated by nodes to transmit data, receive data, aggregate data and final routing of data to BS. Equation (3.4) gives the total transmission energy dissipated. ETX_i denotes the energy dissipated to transmit data packet from member node i to corresponding CH. Similarly, ERX_j is the energy spent by cluster head j to receive data packets and EDA_j is the energy spent on data aggregation by cluster head j .

$$TED = \sum_{i=1}^{mn} ETX_i + \sum_{j=1}^{nc} (ERX_j + EDA_j) \quad (3.4)$$

Initially a random set of individuals or chromosomes are taken. For these the fitness function value is calculated and the general GA steps are imposed on the chromosomes as given in Algorithm 1. Fitness function is calculated for each chromosome as in Equation (3.5). ff denotes the fitness factor and wt denotes the weight of corresponding fitness factor initially taken to be equal for all the three factors considered.

$$Fitness_Chromo(i) = \sum_i (wt_i * ff_i) \quad (3.5)$$

$$ff_1 = TD \div AMD \quad (3.6)$$

$$ff_2 = TED \div EL \quad (3.7)$$

$$ff_3 = nc \quad (3.8)$$

3.2.1 Energy model

In LEACH and TEEN protocols first order radio model is used. The same model is used here also, described as below:

Total energy dissipated for transmitting an l-bit packet is

$$ETX(l, d) = (E_{cir} + \epsilon_{amp} * d^n) * l \quad (3.9)$$

E_{cir} : energy required to activate electronic circuit

$\epsilon_{amp} * d^n$: energy dissipated for transmission of a bit over distance d

The energy dissipated for receiving a K-bit packet is

$$ERX(l) = E_{cir} * l \quad (3.10)$$

3.2.2 Assumptions

- Base station has the knowledge regarding each sensor node in the field. This includes ID-location of the nodes and distance of node from BS.
- Base Station is located far away from the sensor node field.
- Each sensor node has the knowledge of its own energy and location of the BS.
- Initially, every sensor node possess equal amount of energy.
- The energy dissipated for transmission and reception is similar.
- Sensor nodes are stationary once deployed.

3.2.3 Phases

For every number of transmission of data packets WSN undergoes three phases.

- (a) **Cluster formation phase:** Base station implements the genetic algorithm. It has the information regarding every node in the sensor network. And each network node has an identity (ID) known to BS and the node itself. On computation of clusters and regarding cluster heads through GA, BS transmit the information to respective cluster heads. On that event each cluster head transmits a signal to the corresponding member nodes so that they can send the sensed data for further number of rounds to this cluster head for aggregation. CH percentage is predefined and so initialization done through randomness as given in Equation 3.11.

$$\forall s \in \{1, \dots, S\} \text{ and } \forall n \in \{1, \dots, N\}$$

$$Pval_n^s = \begin{cases} 1 & \text{if Energy}(node_n) > 0 \text{ and } randprob_n \leq p \\ 0 & \text{if Energy}(node_n) > 0 \text{ and } randprob_n > p \\ -1 & \text{otherwise} \end{cases} \quad (3.11)$$

where S is the number of individual solutions and N is number of sensor nodes present in the network.

- (b) **Path formation phase:** In this phase, chain of cluster-heads of all the clusters are formed. Every cluster node does not need to propagate the data packet directly to the BS. A path is formed among cluster heads such that data aggregation could be done along the path and final aggregated data is send to the BS. For this to happen, BS transmits the information regarding the path formation to the extreme cluster head of chain formed. That node sends aggregated data along the chain to next node. And finally data sent through cluster head with utmost capacity. This path evaluation is done at BS based on the Euclidean distance between cluster heads, between BS and

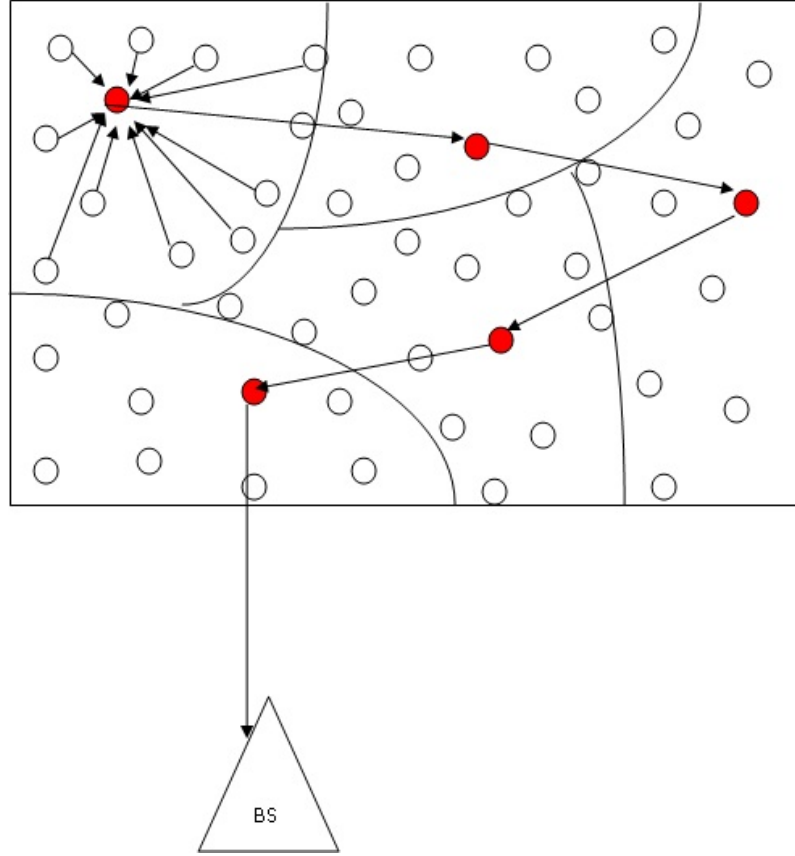


Figure 3.1: Path Formation Phase

every cluster head and the energy left with them. Figure 3.1 depicts the path among the cluster heads towards the BS.

- (c) **Data transmission phase:** Data transmitted only when the two conditions are satisfied: Firstly, the sensed value should be greater than the user adjustable, predefined value termed as hard threshold (HT). And secondly, when difference between latter sensed value and former sensed value is greater than another user adjustable, predefined value termed as soft threshold (ST). In case of data with utmost importance i.e., abrupt increase in sensed parameter, node directly sends the data to the cluster head and then gets propagated to the BS.

3.3 Summary

This chapter gave an in-depth explanation of GA based method used for cluster formation and the fitness function factors used. It gave formulations regarding the methodology and step-wise phases involved in the protocol. The next chapter gives a quantified evaluation of method used in this chapter through simulation results.

Chapter 4

Simulation Result

4.1 Analysis of GA-TSRP

To evaluate the performance of given method, implementation is done using MATLAB and compared with LEACH, TEEN and ERP. Parameters are given in Table 4.1. A 100*100 network field is considered. Also, 100 nodes are taken for simulation. All the nodes are stationary once deployed. Base station is located far from the sensor field. Each node has an initial energy of 0.25 J.

Performance evaluation is done with parameters like energy dissipation, residual energy and number of nodes alive for the comparison of LEACH, TEEN, ERP and the proposed model GA-TSRP.

E_{cir} is the energy dissipated on transmission and reception of per bit data packet. It denotes the energy consumed through the electronic circuit. E_{DA} is the data aggregation energy per bit per message. ϵ_{fs} and ϵ_{mp} are parameters depending on sensitivity (intelligence) of sensor receiver and noise shape which depends on the distance between transmitter and receiver of a message. If the distance is less than a the shortest crossover distance d_o as in Equation 4.1, free space model is used; otherwise, the multi-path model is used.

Table 4.1: Parameter settings for simulation

Parameter	Value
<i>NetworkSize</i>	100*100
<i>Numberofnodes(n)</i>	100
<i>BSLocation</i>	(200, 150)
E_{cir}	50 nJ/bit
E_{DA}	5 nJ/bit/message
ϵ_{fs}	10 pJ/bit/m ²
ϵ_{mp}	0.0013 pJ/bit/m ⁴
E_o	0.25 J
K	4000 bits

$$d_o = \sqrt{\epsilon_{fs}/\epsilon_{mp}} \quad (4.1)$$

Graph plotted in Figure 4.1 shows the nodes alive vs number of rounds for GA-TSRP, ERP, TEEN and LEACH. It is clear from the figure that nodes are alive for greater extent of rounds in case of GA-TSRP than the others. The steep downfall of the number of nodes starts from around round number 900, 800, 550, 450 for GA-TSRP, ERP, TEEN and LEACH respectively.

Graph in Figure 4.2 displays total energy left in the sensor network with increasing round numbers. LEACH tend to show steep decrease in the residual energy. Whereas, TEEN and ERP has better record of energy left. GA-TSRP shows very little steepness for considerable number of rounds proving that uniform usage of node energy done for a longer period of rounds.

Graph plotted in Figure 4.3 shows the total energy dissipation (cumulative) over the rounds for all the four approaches. This shows that in GA-TSRP nodes dissipate much uniformly than others for a longer period of time (rounds). Reasons for this increase in network life-time could be stated as below:

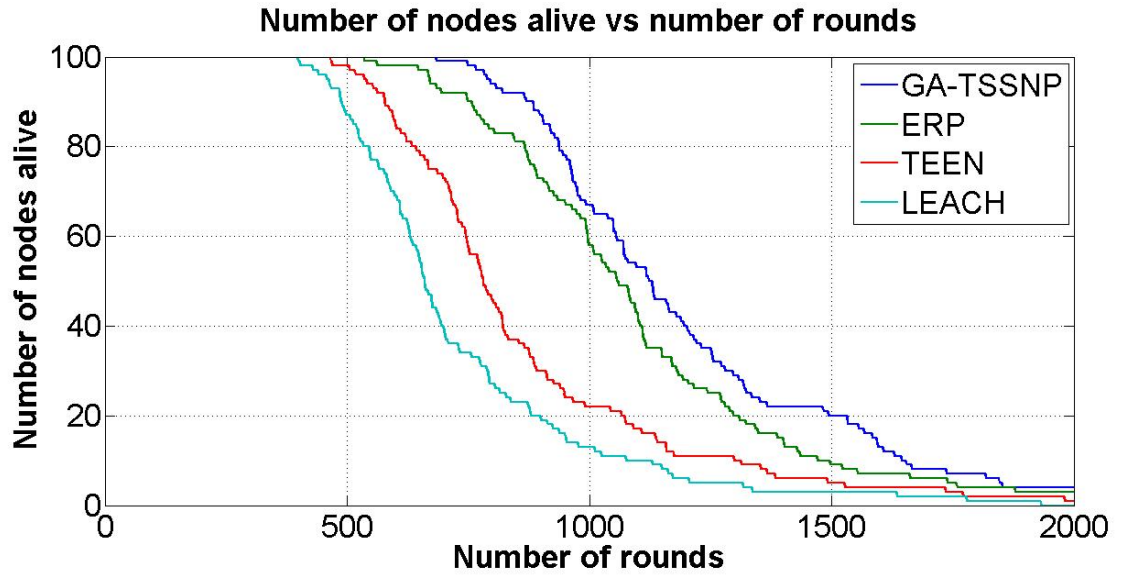


Figure 4.1: Comparison based on nodes alive in network over the rounds

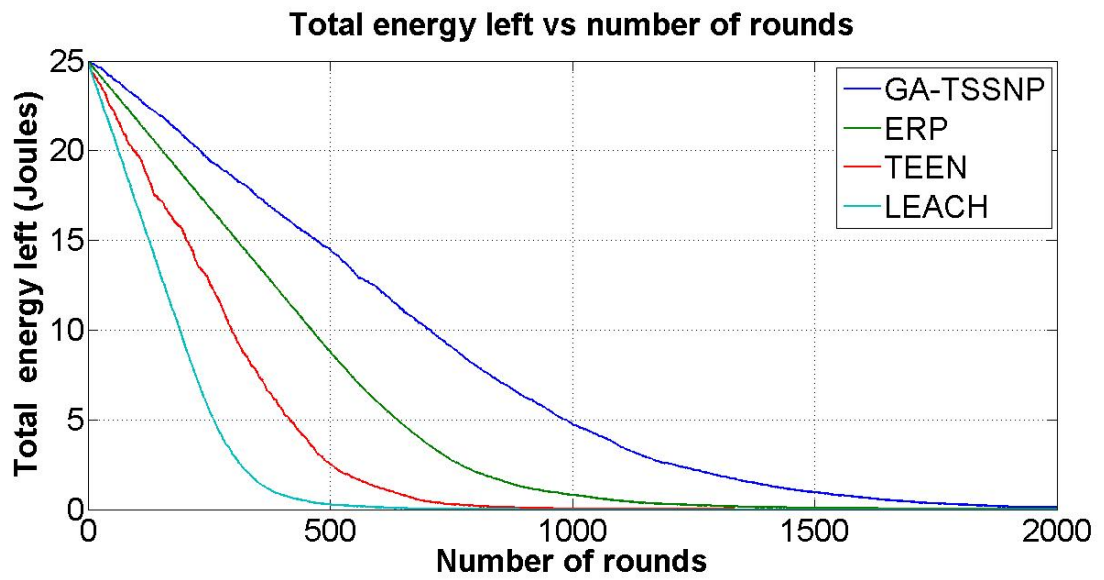


Figure 4.2: Comparison based on energy left in network over the rounds

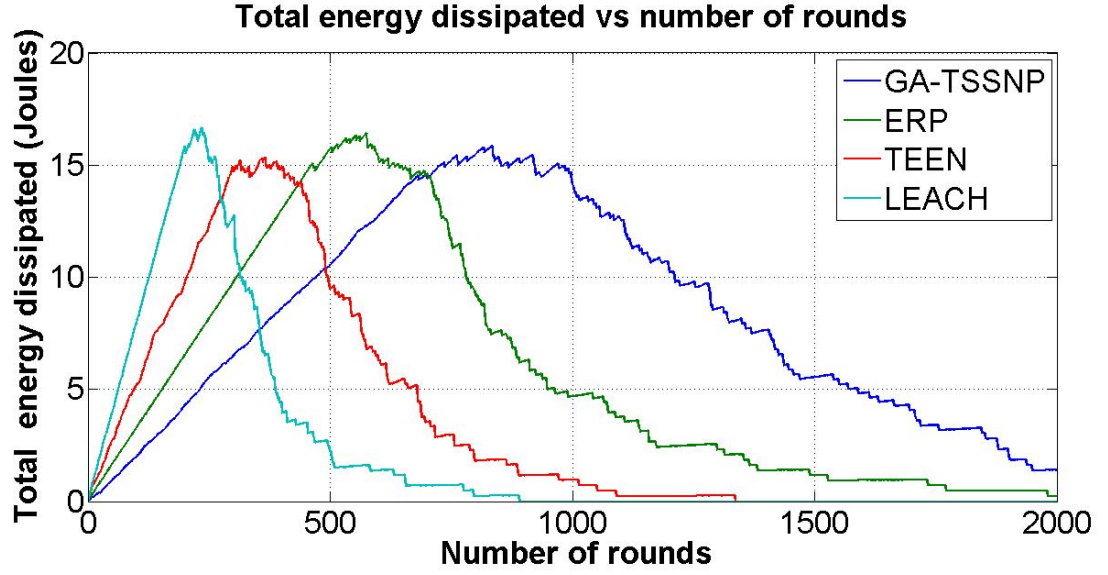


Figure 4.3: Comparison based on cumulative energy dissipation over the rounds

- Fitness function takes into consideration the residual energy present with nodes which will be selected as CHs in any iteration further. It also, takes into consideration both the transmission distance factor relating to cluster heads and member nodes.
- All the cluster heads and especially ones at the extremes of the network does not have to send the aggregated data packets directly to the BS. Instead, a chain is formed among the cluster heads, every time there needs to be a transmission such that only the cluster-heads qualifying on both distance and residual energy basis becomes the end conveyor to BS. In LEACH, TEEN and ERP all the cluster heads send the sensed and aggregated data directly to the BS.
- Cluster head does not depend totally on the threshold value and chances, as in LEACH and TEEN but also the energy left of the node after the transmission in previous rounds.
- Data transmission done on change in attribute value above the threshold.

4.2 Summary

This chapter gave an analytical evaluation of proposed method thus, proving the efficiency achieved with respect to network lifetime and energy associated. The next chapter gives an overview of possible future works related to the method analyzed.

Chapter 5

Conclusion and Future Work

Simulations show that GA-TSRP is efficient than the former protocols with respect to stability period and substantial energy dissipation. These are the outcome of addition of transmission distance and residual energy factors into the fitness function of GA for cluster formation. And also, the node capability based route formation among the cluster heads to BS. Effect of variant weight(wt) on fitness function(ff) can be further scrutinized to give more accurate fitness function. And thus, determining which factor plays a role to which extent so as to determine energy efficient solution. Clear picture can be drawn when worked upon wireless simulation environment by considering other sensor node activities and shortcomings too. Thus, future work can be done regarding testing the proposed method in a network based simulator and deduce weight's importance in fitness function.

Bibliography

- [1] Mohamed Guerroumi, Al-Sakib Khan Pathan, Nadjib Badache, and Samira Moussaoui. Strengths and weaknesses of prominent data dissemination techniques in wireless sensor networks. *International Journal of Communication Networks and Information Security (IJCNIS)*, 5(3), 2013.
- [2] Jamal N Al-Karaki and Ahmed E Kamal. Routing techniques in wireless sensor networks: a survey. *Wireless communications, IEEE*, 11(6):6–28, 2004.
- [3] Wendi Rabiner Heinzelman Anantha Chandrakasan. Energy-efficient communication protocol for wireless microsensor networks. 2000.
- [4] Sajid Hussain, Abdul Wasey Matin, and Obidul Islam. Genetic algorithm for hierarchical wireless sensor networks. *Journal of Networks*, 2(5):87–97, 2007.
- [5] A Attea Baraa and Enan A Khalil. A new evolutionary based routing protocol for clustered heterogeneous wireless sensor networks. *Applied Soft Computing*, 12(7):1950–1957, 2012.
- [6] Sajid Hussain and Abdul Wasey Matin. Base station assisted hierarchical cluster-based routing. In *Wireless and Mobile Communications, 2006. ICWMC'06. International Conference on*, pages 9–9. IEEE, 2006.
- [7] Wendi Rabiner Heinzelman, Joanna Kulik, and Hari Balakrishnan. Adaptive protocols for information dissemination in wireless sensor networks. In *Proceedings of the 5th annual ACM/IEEE international conference on Mobile computing and networking*, pages 174–185. ACM, 1999.
- [8] Chalermek Intanagonwiwat, Ramesh Govindan, and Deborah Estrin. Directed diffusion: a scalable and robust communication paradigm for sensor networks. In *Proceedings of the 6th annual international conference on Mobile computing and networking*, pages 56–67. ACM, 2000.
- [9] MJ Handy, Marc Haase, and Dirk Timmermann. Low energy adaptive clustering hierarchy with deterministic cluster-head selection. In *Mobile and Wireless Communications Network, 2002. 4th International Workshop on*, pages 368–372. IEEE, 2002.

- [10] David Braginsky and Deborah Estrin. Rumor routing algorithm for sensor networks. In *Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications*, pages 22–31. ACM, 2002.
- [11] Fan Ye, Haiyun Luo, Jerry Cheng, Songwu Lu, and Lixia Zhang. A two-tier data dissemination model for large-scale wireless sensor networks. In *Proceedings of the 8th annual international conference on Mobile computing and networking*, pages 148–159. ACM, 2002.
- [12] Ming Liu, Jiannong Cao, Guihai Chen, and Xiaomin Wang. An energy-aware routing protocol in wireless sensor networks. *Sensors*, 9(1):445–462, 2009.
- [13] Yong Yao and Johannes Gehrke. The cougar approach to in-network query processing in sensor networks. *ACM Sigmod Record*, 31(3):9–18, 2002.
- [14] Stephanie Lindsey and Cauligi S Raghavendra. Pegasus: Power-efficient gathering in sensor information systems. In *Aerospace conference proceedings, 2002. IEEE*, volume 3, pages 3–1125. IEEE, 2002.
- [15] Arati Manjeshwar and Dharma P Agrawal. Teen: a routing protocol for enhanced efficiency in wireless sensor networks. In *Parallel and Distributed Processing Symposium, International*, volume 3, pages 30189a–30189a. IEEE Computer Society, 2001.
- [16] Arati Manjeshwar and Dharma P Agrawal. Aptein: A hybrid protocol for efficient routing and comprehensive information retrieval in wireless sensor networks. In *Parallel and Distributed Processing Symposium, International*, volume 2, pages 0195b–0195b. IEEE Computer Society, 2002.
- [17] Georgios Smaragdakis, Ibrahim Matta, and Azer Bestavros. Sep: A stable election protocol for clustered heterogeneous wireless sensor networks. Technical report, Boston University Computer Science Department, 2004.
- [18] Femi A Aderohunmu, Jeremiah D Deng, et al. An enhanced stable election protocol (sep) for clustered heterogeneous wsn. *Department of Information Science, University of Otago, New Zealand*, 2010.
- [19] S Faisal, Nadeem Javaid, Akmal Javaid, MA Khan, Safdar Hussain Bouk, and ZA Khan. Z-sep: Zonal-stable election protocol for wireless sensor networks. *arXiv preprint arXiv:1303.5364*, 2013.
- [20] Sajid Hussain, Abdul Wasey Matin, and Obidul Islam. Genetic algorithm for energy efficient clusters in wireless sensor networks. In *ITNG*, pages 147–154, 2007.

- [21] Hari Rangarajan and JJ Garcia-Luna-Aceves. Reliable data delivery in event-driven wireless sensor networks. In *Computers and Communications, 2004. Proceedings. ISCC 2004. Ninth International Symposium on*, volume 1, pages 232–237. IEEE, 2004.
- [22] V Loscri, G Morabito, and S Marano. A two-levels hierarchy for low-energy adaptive clustering hierarchy (tl-leach). In *IEEE Vehicular Technology Conference*, volume 62, page 1809. IEEE; 1999, 2005.
- [23] Deepak Ganesan, Ramesh Govindan, Scott Shenker, and Deborah Estrin. Highly-resilient, energy-efficient multipath routing in wireless sensor networks. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(4):11–25, 2001.
- [24] Ming Liu, Jiannong Cao, Guihai Chen, and Xiaomin Wang. An energy-aware routing protocol in wireless sensor networks. *Sensors*, 9(1):445–462, 2009.